This week:

1. VAUD: this week I modify paper include（对比上周新计划）

   a) Related work 里可视化推理比较多 但是缺少比较，加一些。

   b) 可视化缺少总览和比较 我编程实现了信息面板和统计图的并列比较 下周会添加固定功能

   c) Case1 里的匹配查询需要重新写一下，文章中写的不清楚

   d) sketch-based 不确切，只是 drag and drop

   e) 局限性和未来工作没有 **Ross**

   f) 人手机的数据有 2-5 千米的误差，需要分析对查询结果的影响或者考虑如何解决低准确性的问题。教父来了 咨询了教父 要用粗粒度数据结构查询来说。下周三之前完成

   g) 历史列表设计 在改

   h) 可视化部分新奇点比较薄弱，这部分可以看我之前的一些周报加一些可视化编码。本科同学完成 本周给本科同学布置了任务并教了 d3.Js

   i) 可视化视图的时间轴应该加以改进 同上

   j) 找 case: Find people who come in the evening to the bar area for meeting their friends. Where do they come from? How many of them come by taxi? Such a query would require integration of information from phone location data, social networks of users, and taxi trajectories. Or: When and where people tweet about traffic jams and what are the destinations of the people involved in the traffic jams? 本周改了数据库例子没找下周找

   k) 查询数据集的选取。可完成

   l) case1 的 od 查询有点复杂，这里需要解释，可以在一个查询节点直接 OD 查询。

写作

m) 添加 expert What background are sers expected to have? Would urban planners be willing/able to use it? Wuld they need any training? 试着完成

n) 文章并不是把所有的数据存在一个 stc 可完成

o) 使用 stc 和 可视化设计的动机 **Ross**

2. 为了预备方案，就是 stc 的使用问题，我把数据存进了数据库并大量修改了程序的提取代码
3. 讲课 ppt 列出讲课各节标题并制作 ppt 中 主要讲的内容是关于城市数据的 VAUD 的程序设计和实现方法技巧 之前介绍一下使用的城市数据特点和重要性。后面说一下城市数据的其他的案例。

Next week：

This week I focus on the programing. Next week, I will focus on paper and finish it.

a) Complete the revision of the paper(未来工作改完了 加上 expert evaluation 和动机三点 需要 ross)
b) Complete PPT

对比 review 又整理了一下修改计划

附录：

Contribution to the field of Visual Analytics

Aiming at interesting patterns that can be derived from heterogeneous urban data, the authors describe the integration of different visualization techniques, supported by interaction capabilities to create an analytical discourse. This clearly matches the core of visual analytics and represents the major contribution of the paper.

Expertise

3 (Knowledgeable)

The Review

The authors present VAUD, a visual analytics tool to support the exploration and comprehension of heterogeneous urban data. Along with the characterization of the proposed query model and a more detailed system description, three case studies (about a lost iphone, traffic jam and human daily life respectively) are illustrated in order to outline the

functionalities of the tool and the advantages coming from its adoption.
A part from small typos (e.g. Fig 4, Cetral Square-> Central Square), the paper is well written and easy to read, and the topics it deal with clearly match the focuses of the conference.
Three major issues should be addressed before the next review phase:

1)Concerning the related works, even if the number of given references is reasonable, the comparison between VAUD and the current state of the art appears a bit weak (only for the visual reasoning). Then, it should be extended to cover the other subsections of the related work.

2)Though it is not directly depending on the authors, the use of mobile phone location data having an accuracy till 5000 meters may arise questions about the validity of the obtained results. Then, the authors should explain how/to what extent this affects their analysis, whether they plan the possibility (for the user) to exclude data having low accuracy.
如何降低误差

3)Probably one or two more detailed and concrete case study/ies (rather than the three "mini" cases shown) would improve the comprehension and explain better the potentialities offered by VAUD

Minor remarks:
- check typos
- Fig.3: I'd suggest to change the term "queried results" into "query results". Though not wrong, it is not so commonly used
- Fig.7: A small icon for "which"/"what" would help

Main Recommendation

Unacceptable for presentation

The Summary Review

The reviewers believe that the paper has potentialities, but still shows several limitations such as:

- limited novelty concerning the visual part and limited interest for the venue (R1,R2,R3)
可视化设计
- weak motivations related to the choice of e.g. STC, visual design (R1, R2)
动机 stc 和 visual design

- weak case studies in term of visualization and content/utility(R1,R2,R3)
- expert feedback (R4)

专家意见

Then, as the improvements required overcome clearly the given revision cycle, we propose to reject the paper, but invite the authors to further work in this direction, as their research would be a valuable contribution to the community.

---------- Submission 262, Review 1 ----------

Overall Rating

3.5

Contribution to the field of Visual Analytics

This is an Application paper describing a tool that allows heterogeneous urban data to be flexibly queried and explored. The paper's contributions are the visual query model and presented visual analysis tool. The paper is well organised and well-written and the tool itself is really impressive – assembles a diverse and large set of data. It also tackles the types of analysis queries that frequently occur in applied, exploratory spatial-temporal analysis: supporting queries of where, when and what occurred and stores some analytic history (provenance) when dynamically building queries. I'm not able to make a judgement on its novelty in terms of data structure – using a spatio-temporal index to speed-up queries. The combination of datasets and queries seems familiar – although for this category of paper I don't think this is a problem. So I do think the work is very impressive and relevant and of interest to the community.

A weakness, especially for a visual analysis tool, is the emphasis in the paper on structural overview and comparison – and a demonstration of how visual overviews and comparison supports/helps guide data analysis. It is great that the authors provide three example case studies that are distinct. However, a question when considering and evaluating a visual tool more generally is: could I have done this without the visualizations? Did the visuals help inform the inquiry? I think there were some opportunities that were missed here. Although the tool provides a user-friendly means of constructing potentially complicated database queries, I do not see the visualizations particularly informing the nature of these queries – one can imagine making the same set of queries independent of the discussed visual summaries (details below).

可视化的作用　可视化纵览和比较

The review/judgement is therefore positive, but it is a shame that there were some

opportunities missed in supporting visual overview and comparison. In the detailed notes are examples of where opportunities are missed in terms of visual summary/comparison – and suggestions for how,

within the review cycle, the authors might elaborate on these in the text. I hope the authors are not discouraged from these comments as this is really impressive work and a lot of care and effort has been taken with the paper and tool. Additionally, other reviewers may see the relative importance of the visual aspects slightly differently -- and a contribution around the importance of the querying environment in particular that I have missed.

Expertise

3 (Knowledgeable)

The Review Detailed feedback.

As above, the paper is well organised and well written – and the tool itself is both impressive and relevant to the discipline. My only concern is with the incisiveness of the visual summaries. Below are notes around the case studies where I'd expect the tool to provide greater visual support/summary. I've tried to highlight aspects of the visual analysis that do support useful summary/comparison and that might be emphasised/elaborated upon in a paper revision. Within the timescales of the revision, it is unreasonable to expect substantive changes to visual views in the tool. The authors may wish to consider a Future Work section elaborating on how views could be further developed to support summary and comparison.

Also see Walker et al. 2013 -- http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6634110. This seems to summarise data that are broadly similar to those appearing in this work, but provides simple visual summaries around the distribution of phenomena that are shown to be instructive and that help inform inquiry.

// -------- Case study 1 ----------

In case study 1, an analyst wishes to support the retrieval of lost property and defines a search for "lost" amongst blog articles. This returns 88 items. From these the analyst selects an entry of interest and uses information in the blog post (a phone lost during a taxi ride) to make a spatiotemporal query on the taxi dataset. The process of making these queries to the dbase is demonstrated in the supporting video. The tool is clearly impressive, supporting text searching and rapid querying. Along with the queries is a map view of geo-tagged blog posts and relevant taxi trajectories. However, these visual summaries do not seem

fundamental to the data analysis/inquiry. One can imagine a scenario in which visual summaries of the text used in a collection of blog posts of interest might help determine priorities for analysis. Frequency distributions around the words preceding and succeeding the word "lost", and represented against some data-driven expectation, might help establish common themes and prioritise/triage certain reported cases. A geo-spatial summary of these posts might reveal further discriminating structure [note: these suggestions are just indicative and shouldn't necessarily be considered within the review cycle].
数据可视化的分析需要一些统计等，词频等

There is a sense that the views and authored examples here fail to show these more complex structures – I don't see the visual summaries as determining/helping to shape the analytic inquiry; the same set of queries could have been imagined without the demonstrated visualizations. However, the final analysis task for case study 1 – comparing trajectories derived from mobile phone and taxi – is a task uniquely supported by the visualization -- so the authors should emphasise this. Whilst a summary statistic might be used for this comparison (negating the need for visualization), one can imagine idiosyncratic features associated with a trajectory not accounted for in a summary statistic, but that might be uniquely identified through visual inspection.

\\ ---------- Case study 2 ----------
In case study 2, a traffic jam is analysed. The authors query on a set of conditions relevant to traffic jams: low speed for a single day. It is great that the authors use a visual representation of taxi trajectories combined with a heat map of average speeds to identify streets with large numbers of trajectories of low speed. However, one might imagine a set of visual summaries that enable more rich/detailed structure here: e.g. small multiple maps perhaps summarising particular parts of the day. Later, the authors provide summaries of land-use/points of interest (POIs) around two selected streets with low average speeds and compare the modal POI for each. Again, one might imagine more expressive summaries used to characterise/compare this distribution of land use. Why not represent land use directly on the map view and combine with a summary of average travel speeds for specific road sections? Again, I don't suggest this as a substantive edit within the review cycle – but the paper's supporting text could be adapted to elaborate more on these possibilities for visual comparison.
小图　　热力图在地图上显示
\\ -------- Case study 3 -------

Case study 3 is about comparison: the authors wish to compare the lives

of residents from two contrasting villages – the villages differ based on their house prices. The authors identify two residents from villages by checking geo-tagged phone data at night (a plausible proxy for home location). They then gradually introduce and compare more information – e.g. social networks, daytime visit locations. Again, this case study nicely demonstrates how database queries can be made, but if the aim was genuinely to compare the lives of residents from villages of contrasting house prices, then surely it would make more sense to identify all individuals in the dataset that live in those areas and study the visiting locations and social network characteristics of these two groups -- so visually depicting group behaviour. It is of course possible that any identified differences between two individuals might relate to factors other than where those individuals live – analysis and comparison of variation within and between groups is therefore necessary and visualization might support this analysis.
第三个例子能不能设计比较两组人，分析群体行为

Some small typos/ syntax errors

- p2, bullet 1: and deduction from a multiple, data sources
- p2, final paragraph of Section 1: “and case studies provide serve as an exemplar”
- p6, Fig 8 caption. “(a) A retrieving operator. (b) A reasoning operator.” – in text b) is called extraction operation.”
- p6, Section 4.2, paragraph 1: “In the scene view, The road network”
- p6, Section 4.3, point 1: “example,to query”

---------- Submission 262, Review 2 ----------

Overall Rating 2.5

Contribution to the field of Visual Analytics

The contribution of the paper is a visual query interface for performing queries over several pre-selected datasets. It allows users without knowledge of database query languages to specify simple queries on one of the datasets and use the results for making further queries on other datasets. Thereby the user creates query chains, which are represented visually by a linear flow chart. The analytical capabilities of the system are quite limited: the system supports searching for elementary facts but not larger scale analyses over space and time. The query UI and

visualization of query chains are not novel. The paper can be mostly considered as a demonstration of the possibility of supporting users in extracting and linking data from several heterogeneous datasets in a relatively simple way. I am hesitant if this contribution is sufficient for the VAST conference.

Expertise 4 (Expert)

The Review

The strength of the proposed system is that a user can build quite complex queries from elementary queries which are relatively easy to specify using a direct interaction UI. The authors claim that the system supports sketch-based user interactions, but this is a wrong use of the term: the user does not make any sketches (i.e., drawings) but interacts mainly by dragging and dropping.

The interactive visual UI for querying is not by itself a novelty; previous works in this area are referred to in the paper. The novelty is in supporting querying across several datasets, which has not been previously done in a visual way. However, the system described does this for several pre-selected datasets. It does not seem to be flexibly adaptable to another combination of datasets but is designed to represent particular data components in specific pre-selected ways.

What is definitely missing for a proper support of queries across different datasets is a representation of the structures of these datasets and the possible links, i.e., which fields in one dataset correspond to which fields in another dataset. In the presented use scenarios, it is assumed that the user knows very well what is in each dataset and how they may be linked. Still, explicit representation of the database structures and possible links would be a helpful reminder even for an expert user.
数据之间的联系
I wonder concerning the purposes of bringing all data to the space-time cube representation (STC), which requires much memory for storage while being very sparsely filled with data items. I see no real need in such a representation since queries involving two or more datasets are built from several atomic queries, and each atomic query deals with only one dataset. Hence, it would be quite easy to use results of one atomic query for creating the next atomic query to another dataset. The queries can be straightforwardly translated to standard database queries, e.g., in SQL, and performed using standard database functionality. So, I do not see any benefits in using the STC representation. Anyway, not all datasets can be

transformed to this format. POI data have no temporal information and therefore do not fit in an STC. The social network data are also hard to represent in an STC. Hence, in my understanding, the transformation to STC is neither needed nor sufficient.

Concerning the visual query interface, I see many possibilities for improvement. Thus, the way in which OD queries are specified is complicated and not intuitive. Instead of separately extracting a set of trips by origin and a set of trips by destination and then making an intersection of these, it might be more straightforward and easier for the user to specify both the origin and destination in one query condition. The possibilities for querying by time are now limited to selecting intervals along a time line. This might be extended by controls corresponding to the cyclic time model. An example is Time Wheel:

@proceedings {6051,
title = {A Graphical User Interface for the Integration of Time into GIS},
journal = {1997 American Congress of Surveying and Mapping Annual Convention and Exhibition},
year = {1997},
pages = {182-189},
address = {Seattle. WA},
url = {http://www.geovista.psu.edu/publications/ACSM97/index.html},
author = {Donna J Peuquet and Edsall, Robert M.,}
}

Still, I see the main limitation of the system in that it supports answering only elementary questions, in terms of Bertin, and is not suitable for analysis on intermediate and overall levels. The example use cases described in the paper are just toy examples, which make an impression of being driven by mere curiosity rather than by real analysis needs. There are also aspects that raise doubts. In case 1, is the user identifier in the phone call database the same as in the social media? This is not usual, and establishing correspondences between user IDs in different databases is not a trivial task. In example 2, the POIs in the traffic congested areas would be immediately visible if the result of the query step 1 were shown on top of an appropriate geographic background, such as map tiles from Open Street Map. In this case, the user would not need an additional query for finding the POIs. In all examples, it is unclear how the street view contributes to answering the questions. In example 3, a pair of persons from a wealthy and a poor villages are arbitrarily selected. From comparing their social networks, it is not valid to draw a general conclusion that people in the wealthy village

have more social connections than people in the poor village. Actually, almost the whole example 3 consists of arbitrary selections, which is typically not a useful approach in real-world data analyses and problem solving.

一些关于例子的建议

I believe that the system has potential for enhancing its analysis-supporting capabilities. Thus, by combining data from multiple sources, it is possible to answer sophisticated analytical queries like: Find people who come in the evening to the bar area for meeting their friends. Where do they come from? How many of them come by taxi? Such a query would require integration of information from phone location data, social networks of users, and taxi trajectories. Or: When and where people tweet about traffic jams and what are the destinations of the people involved in the traffic jams? Such queries require, first, some extensions of the query UI (possibly, with a visual equivalent of GROUP BY operations available in SQL) and, second, data aggregations and overall views over space and time. Furthermore, it would be good to be able to select all areas or time intervals with aggregate values above or below user-chosen thresholds and use the selection as a query condition for following queries.

Not only such extensions would enhance the analytical capabilities of the system, but focusing on the intermediate and overall level questions, in Bertin's terms, instead of accessing personal data of individuals, would reduce the risk of compromising personal privacy.

My general conclusion is that the current system is too weak for supporting truly visual analytic activities beyond answering elementary questions by retrieving and showing individual data items. There is potential for strengthening, but the current paper may be of limited interest to the audience.

---------- Submission 262, Review 4 ----------

Reviewer: external

Overall Rating

4: Good contribution

Contribution to the field of Visual Analytics

This work proposes a visual system to explore the heterogeneous urban data. As far as I know, it is one of the very first several works fusing

the spatial-temporal-social data together. This work presents a powerful visual exploration with smooth interactions, which supports the cross-domain query as well as correlation in the urban data.

Expertise

3 (Knowledgeable)

The Review

This work proposes a visual exploring method for the heterogeneous urban data, covering the cyber-physical-social space. The approach supports to query for cross-domain correlation from multiple data sources via sketch-based user interactions. The query model provides a power exploration of the objects from four distinct attributes, i.e. which, when, where and what the object it is.

The dataset and problem of this work is very interesting. As far as I know, it is one of the very first several work correlating the spatial-temporal-social data together. The fusion of multiple urban data sources is one of the biggest challenges in urban data analysis, this work provides a powerful visual exploration via smooth interactions. The query model is well structured which fully supports the cross-domain query.

The interface of the system is very neat and effective. The paper is well-structured and easy to following. Figures in the paper are nice and well designed. However, there are some points suggested to be revised:

First, the details of design, techniques are clarified in the paper, but the evaluation of the system is expected with more clarifications. Now it covers three case studies. As mentioned in Section 5, Paragraph 1#, this approach is proposed based on the cooperation with domain exports in urban planning and public security filed. Besides to the cases, the feedbacks from those exports will demonstrate the effectiveness and usability of the system better.

Second, the exploration interface seems to use the Flow metaphor, which I notice the branching queries in the video. However, there is no text explanation about it. Although Flow metaphor is not new in query, I think it should be consistent between the paper and its supplementary material.

例子文字修改

Third, in Case 3 (Section 5.3), the explanation about the relationship

between surrounding POIs and high/low price village is confusing. The meaning of “contained trajectories” is not clear. The two parts before and after the phase “In contrast” compares the “individuals living in expensive apartments” with “persons in high-price village”, which refers to the same group people.

Finally, as one of the first researches in fusing multiple source data, more discussion about the work, its limitations and future work will complete the work better.

As a summary, although there are something to be revised, I locate this work as a potential one in VAST field which will encourage and inspire the researching in analysis of multiple source urban data.

---------- Submission 262, Review 5 ----------

3: Sufficient contribution

Contribution to the field of Visual Analytics

The main contributions of this work are 1) a visual query model supporting reasoning on multiple heterogeneous sources of spatio-temporal data; 2) a visual analytics framework supporting the visualization, querying and reasoning on such data based on the aforementioned model. The usefulness of the model is demonstrated through two case studies requiring the cross-analysis of multiple data sources.

Expertise

2 (Passing Knowledge)

The Review

The main contributions of this work are: a visual querying model and its implementation in a visual analytics framework supporting the analysis of multiple heterogeneous sources of urban data, which is a problem relevant to the VAST conference. The paper makes a convincing description of the challenges met in the interactive analysis of such complex data and demonstrates its usefulness in a credible way through two well-developed use cases. I also appreciate the fact that the system was developed in connection with real users in the field of urban planning and that the featured case studies use real datasets.

The main weakness of this paper is the absence of a user study. In its current form, the paper does not give any indication on the usability of such an elaborate querying model by the target users. What background are users expected to have? Would urban planners be willing/able to use it? Would they need any training? The paper would be much stronger, had it

covered these considerations. Given that there is ample space left, the authors could have easily added such a discussion. I am therefore not inclined to give a higher score.

The paper reads very well, is well-organized and makes an adequate use of figures to illustrate its point. It is adequately referenced and clearly positioned in the context of previous work. The supplementary material was helpful in appreciating the maturity of the system and its typical use.

The paper has a few typos that can easily be fixed, and a strange hyphenation pattern (producing single "s" letters at the beginning of lines at multiple places).